ABSTRACT

This invention relates to a text-to-speech conversion system (TTS) for interlocking with multimedia and a method for organizing input data of the same. A conventional TTS is in situation of the only for the synthesis of speech from the inputted text. In addition, by a prior organization, since it is impossible to presume from only the text the information required when moving picture is to be dubbed by use of TTS or when the natural interlock between the synthesized speech and multimedia such as animation is to be implemented, there is no method to realize these function. Furthermore, there is also no result of the studies on use of additional data for enhancement of the natural in the synthesized speech and organization of these data. Therefore, an object of the present invention is to provide a text-to-speech conversion system (TTS) for interlocking multimedia and a method for organizing input data of the same for enhancing the natural of synthesized speech and accomplishing the synchronization of multimedia with TTS by defining additional prosody information, the information required to interlock TTS with multimedia, and interface between these information and TTS for use in the production of the synthesized speech. According to the present invention, a foreign movie can be dubbed in Korean by implementing the synchronization of the synthesized speech with the moving picture by way of the direct use of text information and lip-shape information which is presumed by the

analysis of actual speech data and lip-shape in the moving picture for the production of the synthesized speech. Still furthermore, the present invention is applicable to a variety of field such as communication service, office automation, education and so on by making the synchronization between the picture information and the TTS in the multimedia environment possible.

conversion system (TTS) for interlocking with multimedia comprising the steps of:

classifying multimedia input information organized for enhancing the natural of synthesized speech and implementing the synchronization of multimedia with TTS into text, prosody, the information on synchronization with moving picture, lip-shape, and individual property information in a multimedia information input unit;

distributing the information classified in the multimedia information input in a data distributor by each media, based on respective information;

converting text distributed in the data distributor by each media into phoneme stream, presuming prosody information and symbolizing the information in a language processor; calculating a value of prosody control parameter other than prosody control parameter included in multimedia information in a prosody processor;

adjusting the duration every each phoneme in a synchronization adjustor so that processing result in the prosody processor may be synchronized with a picture signal according to input of the synchronization information;

producing the synchronized speech in a signal processor using the prosody information from the data distributor by each media, the processing result in the synchronization adjustor, and a synthesis unit database; and

outputting the picture information distributed by the data distributor by each media onto a screen in a picture output

apparatus.

3. The method according to claim 2, wherein said organized multimedia information is comprised of text information, prosody information, information synchronized with moving picture, lip-shape and individuality information.

4. The method according to claim 3, wherein said prosody information is comprised of the number of phoneme, phoneme stream information, duration time of each phoneme, pitch pattern of the phoneme and energy pattern of the phoneme.

5. The method according to claim 4, wherein said duration of the phoneme is indicative of a value of pitch at beginning point, middle point, and end point within the phoneme.

6. The method according to claim 4, wherein said energy pattern of the phoneme is indicative of a value of energy in decibel at beginning point, mid point and end point within phoneme.

7. The method according to claim 2, wherein said synchronization information is comprised of text, lip-shape, location information with moving picture, and the duration information.

8. The method according to claim 2, wherein said synchronization information is composed of a beginning point, duration and delay time information of starting point, and duration of each phoneme

is controlled by said synchronization information.

9. The method according to claim 2, wherein said synchronization information is composed of a duration of the beginning point of a sentence and a duration information of starting point, and duration of each phoneme is controlled by forecast lip-shape considered an articulation manner of the phoneme and articulation control,

lip-shape within the synchronization and duration information composed of said synchronization information.

10. The method according to claim 2, wherein said synchronized speech is produced by an information of beginning point and end point of each phoneme related with speech signal and an information of phoneme.

11. The method according to claim 2, wherein said synchronized speech is produced by a numeralization of distance(extent of opening) between upper lip and low lip, distance(extent of width) between left and right end points of lip, and extent of projecting of lip and the lip-shape quantized and normalized pattern depended on articulation location and articulation manner of the phoneme on the basis of pattern with high discriminative property.

12. The method according to claim 2, wherein said transmission method of multimedia information comprising the steps of:

converting a prosody information existed in the multimedia information into a data structure capable of utilizing in the signal processor;

transmitting the converted prosody information to the prosody and the synchronization adjustor;

converting the prosody information outputed from the prosody and the synchronization adjustor to a data structure capable of utilizing in the synthesis unit database and the prosody processor within the TTS if the prosody information is included in said multimedia input information;

transmitting then to the synthesis unit database and the prosody processor if the individual property information is included in said multimedia input information.

add a3